

# PREDECIR LA MOVILIDAD DE LOS EMPLEADOS.

---

Tras cursar un Bootcamp gratuito de Data Science.



# OBJETIVO

---

- Predecir si el trabajador buscará un nuevo puesto de trabajo o trabajará para la empresa cuando finalice la formación.
- Problema de clasificación.



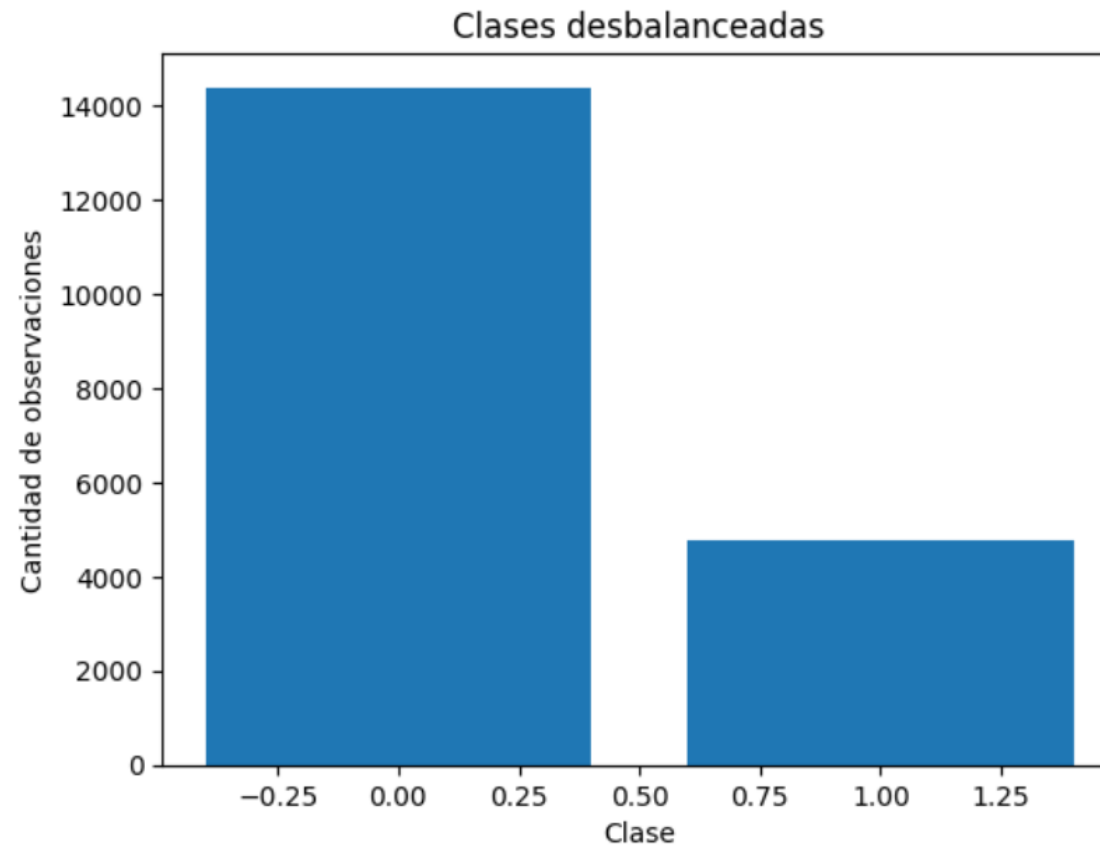
# PROCESO DE LIMPIEZA Y TRANSFORMACIÓN DE LOS DATOS

- Clases desbalanceadas:

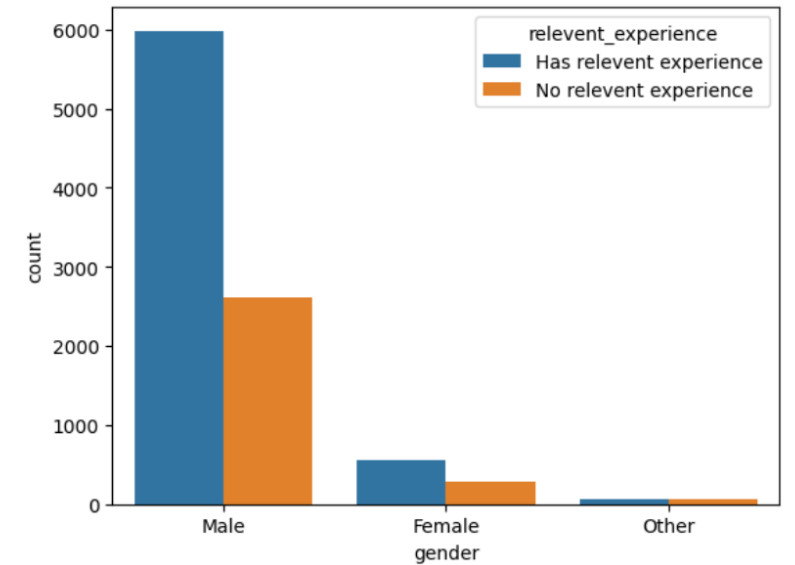
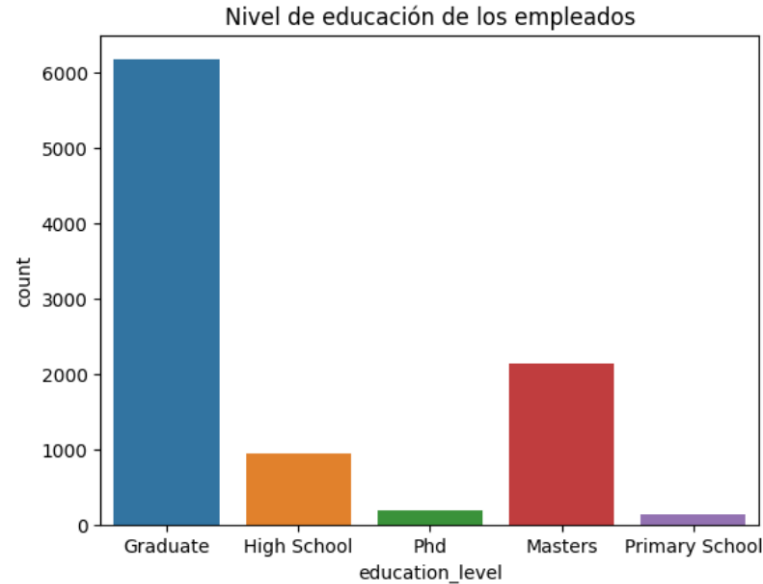
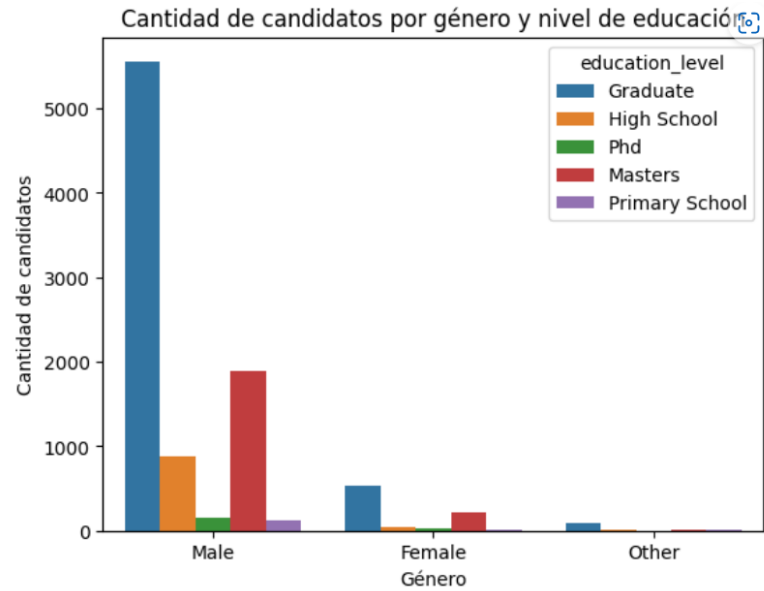
- 14.381 clase 0: No buscan  
cambio de trabajo.

- 4.777 clase 1: Buscan  
cambio de trabajo.

- Para solucionar este problema, vamos a realizar la técnica “Undersampling” para quedarnos con el mismo número de muestras de ambas clases y poder entrenar el modelo con las clases balanceadas.

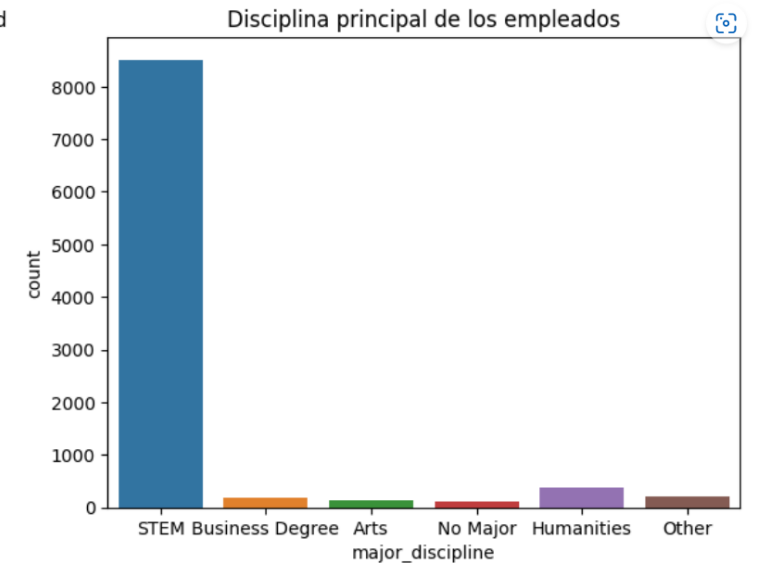
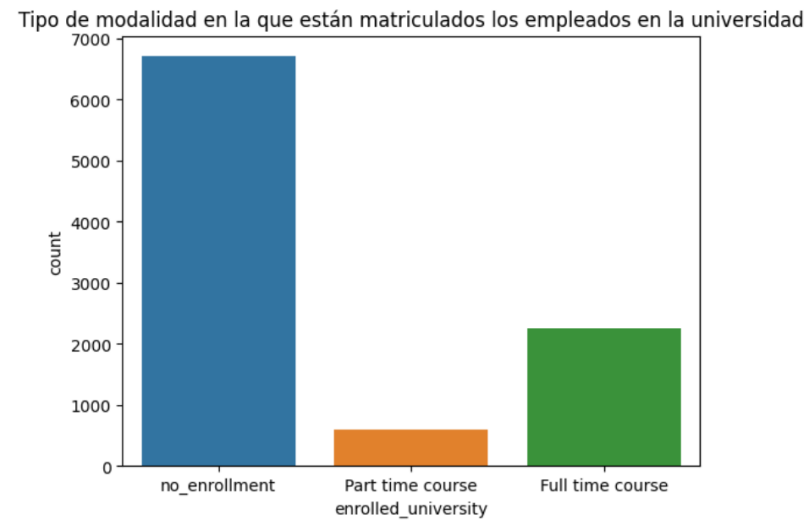
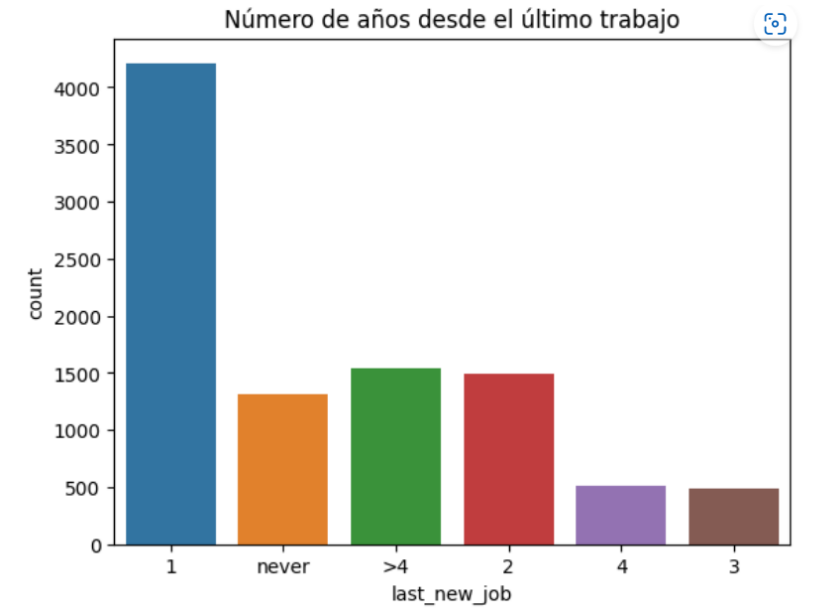
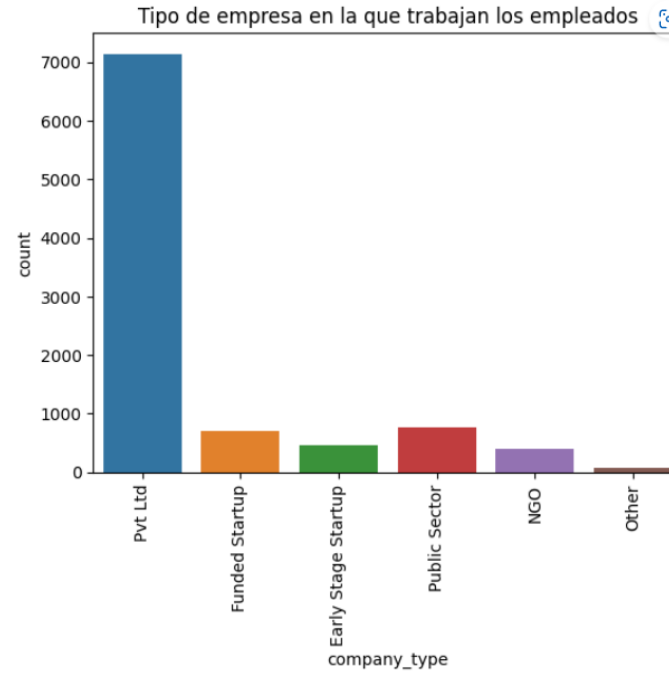


# ANÁLISIS EXPLORATORIO DE LOS DATOS



# ANÁLISIS EXPLORATORIO DE LOS DATOS

---






# MODELOS UTILIZADOS

- **1-Random Forest Classifier**
- **2- AdaBoost Classifier**
- **3-XGBoost Classifier**
- **4-LGBM Classifier**
- **5- Catboost Classifier**
- **6- Support Vector Classifier (SVC)**
- **7- Red Neuronal.**
- **8-EasyEnsemble Classifier**
- **9-Balanced Random Forest Classifier**
- **10-RUSBoost Classifier**





# MÉTRICA SELECCIONADA Y RESULTADOS OBTENIDOS

- La métrica más importante para nuestro problema de negocio es la PRECISIÓN, ya que mide la proporción de casos positivos que han sido correctamente clasificados como positivos, es decir, la proporción de candidatos que realmente quieren trabajar en la empresa después de formarse que han sido identificados correctamente como tales por el modelo.
- Tras una primera valoración, la métricas que mejor me salen son:
  - 1)- Catboost con una precisión de 0,73.
  - 2)-Red Neuronal con una precisión de 0,767
  - 3)-RUSBoost Classifier con una precisión de 0,75



# OPTIMIZAR LOS MODELOS

---

- Optimizo los hiperparámetros de los 3 modelos que mejor resultado me han dado:
  - 1) Catboost: la precision mejora a 0,78.
  - 2) Red Neuronal: la precision se queda igual.
  - 3) RUSBoost Classifier: la precisión se baja a 0,72





# MODELO FINAL SELECCIONADO

---

- He escogido el Catboost Classifier como modelo ganador:
  - Mejor métrica de precision proporciona.
  - El recall tambien es alto, y se puede tomar como medida de apoyo ya que identifica los empleados que realmente quieren trabajar para la empresa y permitirá a la empresa maximizar sus oportunidades de contratación.
  - Es el modelo que menos tarda en entrenar.





## MANTENER Y MEJORAR EL MODELO SELECCIONADO

- Crearé una función de comprobación de métricas que nos permita evaluar la calidad del modelo actualizado en comparación con el modelo anterior y guardaré el modelo entrenado si es mejor.





GRACIAS  
BEATRIZ MARTÍNEZ GARCÍA

---